Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor

Andrew D. Lawrence[#1], Stefanie Frank[#1], Sarah Newnham[1], Matthew J. Lee[1], Ian R. Brown[1], Wei-Feng Xue[1], Michelle L. Rowe[1], Daniel P. Mulvihill[1], Michael B Prentice[2,3], Mark J. Howard[1] and Martin J. Warren[1]

1. School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent, CT2 7NJ, UK
2. Department of Microbiology, University College Cork, Cork, Ireland
3. Department of Pathology, University College Cork, Cork, Ireland

[#] Have contributed equally to this work

*Running title: *Bacterial microcompartment targeting peptides*

To whom correspondence should be addressed:
Prof Martin J. Warren, email: m.j.warren@kent.ac.uk, tel 00 44 (0)1227 824690

Keywords: Metabolic engineering, compartmentalization, propanediol utilization, synthetic biology

Abstract

The targeting of proteins to bacterial microcompartments (BMCs) is mediated by a short peptide sequence of 18 amino acids. Herein, we report the solution structure of the N-terminal targeting peptide (P18) of PduP, the aldehyde dehydrogenase associated with the 1,2-propanediol utilization metabolosome. The solution structure reveals the peptide to have a well defined-helical conformation along its whole length. Saturation transfer difference and transferred NOE NMR has highlighted the observed interaction surface on the peptide with its main interacting protein, PduK, a component of the outer shell of the microcompartment. By tagging both a pyruvate decarboxylase and an alcohol dehydrogenase with targeting peptides it has been possible to direct these enzymes to empty BMCs *in vivo* and to generate an ethanol bioreactor. Purification of the re-designed BMCs reveals that not only do they contain the ethanogenic enzymes but that they are able to transform pyruvate into ethanol efficiently.

Introduction

Bacterial microcompartments (BMCs) are discrete metabolic units found in a range of bacteria that are dedicated to a specific metabolic pathway (1). They consist of a semi-permeable proteinaceous outer layer that encases enzymes associated with a particular process. The best characterized of the BMCs is the carboxysome, which is found in cyanobacteria and some chemoautotrophs (2-7). Here, the enzymes carbonic anhydrase and RuBisCo are retained within the confines of the macromolecular complex to provide an environment for enhanced carbon dioxide fixation. Other examples of the BMCs include the metabolosomes associated with both 1,2-propanediol and ethanolamine utilization (8-13). Of these the former has been the subject of greater investigation and characterization.

The propanediol utilization (*pdu*) operon is composed of 23 genes (11, 14, 15) and encodes largely for proteins that form a supramolecular complex in the form of a metabolosome (13) with a diameter of between 100 and 150 nm. Six of the genes (*pduABJKUT*) encode for shell proteins that form hexameric tiles, which are envisaged to align together to form the facets and edges of the outer casing of the structure (16-19). The vertices of the BMCs are thought to be formed from the pentameric PduN (7, 20). The shell proteins snare the enzymes for 1,2-propanediol metabolism, including the diol dehydratase (PduCDE), and the alcohol and aldehyde dehydrogenases (PduP and Q) (8, 11). The metabolosome also houses enzymes for the repair and reactivation of the diol dehydratase (PduG, H) and its coenzyme (PduO, S), adenosylcobalamin (8, 11). The shell of the metabolosome has to allow the passage of its substrates, cofactors and coenzymes into the bacterial microcompartment as well as the exit of the metabolic products (18, 21, 22). This is likely mediated through the central pores that are formed within the tiles of the shell structure (16, 23). Other proteins are thought to interact with the shell proteins on the external surface of the structure, including PduV, which may help to localize the metabolosome within the cell (24). A summary of the proposed Pdu BMC is given in Figure 1.

We have shown that it is possible to generate recombinant microcompartments in *Escherichia coli*, through the transposition of the whole *pdu* operon from *Citrobacter freundii* to generate fully functional metabolosomes (11). More recently it has also been reported that recombinant carboxysomes can be produced in *E. coli (25)*. Through the coordinated production of just the Pdu shell proteins empty microcompartments can be constructed within the cell (24). A minimal set of shell proteins, PduABJKN, appears to be required for the assembly of empty

metabolosomes although slightly larger units are formed if PduU is included with the other shell proteins. Proteins can be targeted to these empty microcompartments by tagging them to other proteins that are found associated with the BMCs such as PduC, PduD and PduV (24). The potential therefore exists to generate new bioreactors within the molecular scaffold of a bacterial microcompartment (1) (Figure 1).

Key to understanding how BMCs form is the localization of the metabolic enzymes to the inside of the structure. It is not known if the shell of the metabolosome forms around the metabolic enzymes or whether the enzymes are internalized after the initial assembly of the compartment. However, it is known that encapsulation of some of the enzymes is dependent upon the presence of a peptide targeting sequence (26, 27). The targeting sequence appears to be located on either the N- or C-terminal region of the internalized protein (28). However, the structural features associated with these interactions have yet to be determined. For proteins such as PduP and PduD this targeting sequence can be very short, comprising around 18 amino acids (26, 27). Sequence analysis predicts that these targeting sequences are likely to be helical in nature. Herein, we provide the solution structure of the N-terminal 18 amino acids of PduP from *C. freundii* and show how this peptide sequence, together with a similar peptide from PduD, can be used to generate a simple ethanol bioreactor within a Pdu microcompartment shell by targeting the *Zymomonas mobilis* enzymes pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (Adh) to the complex.

Results

PduP interacts with PduK

Previous research has shown that the N-terminal region of PduP is effective in targeting protein to the BMC (24, 27). In order to determine which of the shell proteins interact with PduP we conducted a series of pull-down assays in which PduP was co-produced in *E. coli* along with individual shell proteins harboring an N-terminal poly-histidine tag. Co-purification of PduP alongside a His-tagged shell protein following immobilized metal affinity chromatography (IMAC) is indicative of a protein-protein interaction and the formation of a tight complex. Using this method, we observed that PduP and PduK purified together (Supplementary Figure 1). PduK is known to be an abundant constituent of the shell and its interaction with the P18 peptide (the first 18 amino acids of the N-terminus of PduP) was further investigated.

To determine the binding constant for the interaction between the P18 peptide and PduK we monitored the intrinsic tryptophan fluorescence of PduK, which contains only one tryptophan residue. Titration of P18 into a solution of PduK resulted in quenching of the fluorescence signal from the tryptophan. An equation describing a hyperbolic decay was fitted to the data yielding a K_d of 331 ± 64 nM (Supplementary Figure 2). This titration was further evidence of a strong interaction between P18 and PduK. Moreover, the K_d is in a range that is acquiescent for STD NMR (29).

PduK (156 amino acids) is larger than shell proteins such as PduA, PduJ and PduU as it contains a C-terminal extension that has been speculated to house a Fe-S center (16). To investigate if P18 binds to the BMC domain within PduK or whether it interacts with the C-terminal region of the protein a truncation of the protein was made that consisted only of the first 96 amino acids. This truncated variant of PduK was also found to interact with P18 and with a similar K_d to the full-length protein.

Solution structure of the P18 peptide

The N-terminal region of PduP, which is responsible for the incorporation of the PduP enzyme into the lumen of the microcompartment, is suggested to be helical by both secondary structure prediction and CD spectroscopy (30). Since there is no crystal structure available for this region of PduP or its homologues we decided to solve the solution structure of the synthetic P18 peptide using NMR spectroscopy.

The structure of the P18 peptide (NH_3-MNTSELETLIRNLISEQL) was determined by ^{1}H NMR in both the presence and absence of 30% (v/v) trifluoroethanol (TFE). Chemical shift assignments and through-space structural information were obtained from two-dimensional TOCSY and NOESY NMR experiments (Supplementary Table 1 and Supplementary Figure 3). The observed NOE contacts support the presence of an α-helical conformation of the peptide with NOEs observed between Hα and H_N (i–i+3) as well as Hα and Hβ (i–i+3) in both the presence and absence of TFE (Supplementary Figures 2a and 4 and Supplementary Table 2). Calculated structures demonstrate the formation of a continuous α-helix between residues 3-16 (Supplementary Figure 4). This is despite assignable Hα-H_N and Hα-Hβ i-i+3 NOEs not being continuous across the entire range of the helix due to spectral overlap. The presence of additional side chain i–i+3 NOEs and weak Hα-H_N i–i+4 NOEs confirm the continuity of the helix. It is important to note the helical nature of the peptide in the absence of TFE. The addition of TFE, which is typically used as a stabilizer to aid secondary structure formation in short peptides, reduces the length of the helix (residues 5-15) and increased flexibility of the termini when compared to P18 in the absence of TFE (Supplementary Figure 4). In this case, TFE has acted to destabilize the native structure of the P18 peptide. The dualities of TFE acting as both a stabilizer of secondary structure and also as a denaturant has been studied and documented (31). We suggest that the helical secondary structure element must be central to the process of recognition between the shell proteins and the PduP targeting sequence.

Saturation transfer difference (STD) NMR

The interaction between the P18 peptide and the shell protein PduK was investigated by 1D saturation transfer difference (STD) NMR (Supplementary Figure 5). This technique permits the detection of transient binding of a small molecule to a larger protein or receptor and can be used to determine molecular regions of the ligand that have primary contact and are most likely responsible for recognition and binding to the protein of interest. Due to spectral crowding it was not possible to obtain unambiguous STD values for all protons in the P18 peptide and only resonances that were not overlapping were used in the subsequent calculations and analysis. ^{1}H T_1 and STD NMR contributions were obtained through correlation of signal intensity to specific nuclei from the peptide. STD factors were modified according to the T_1 relaxation rate to provide a quantitative STD (qSTD) contribution (Supplementary Table 3). These values were displayed as percentage transfer compared to the maximum value (Figure 2c) and this data was mapped onto the calculated helical structure of the peptide (in the absence of TFE) to illustrate the observed interaction surface (Figures 2a and b). The residues with maximal STD factors

highlight the center of one face of the helix. A comparison of N- or C-terminal BMC-targeting sequences has highlighted a number of conserved hydrophobic residues (28). There would appear to be an apparent motif consisting of two hydrophobic residues followed by two polar residues followed by a further two hydrophobic residues. This represents the LIRNIL region in the middle of the P18 sequence and moreover the area with the highest STD values.

Transfer NOE

The conformation of the P18 peptide while bound to PduK was investigated by NMR using transferred NOEs (trNOE). The trNOE allows structural information to be gathered on a ligand in its bound state while appearing on the resonances of the free ligand. Transferred NOE experiments were carried out on samples containing 1 mM P18 peptide and 66 µM PduK. With approximately 15-fold excess of ligand and a moderate binding affinity, the peptide is in rapid exchange between free and bound states. NOEs indicative of the free state develop slowly. Thus, NOEs observed in the transferred NOE spectrum, at short mixing times, are indicative of the bound conformation of the peptide.

At short mixing times (100 ms) the observed NOE contacts were almost identical to the control NOESY spectra obtained in the absence of PduK, which indicates that there is no structural change present upon binding. The major difference was an increased maximal NOE intensity and an increased rate of NOE build-up in the presence of the binding partner PduK (Figure 3). The increase in the rate of NOE build-up and intensity is due to the binding of the P18 peptide to the PduK protein, which enhances the cross relaxation rate due to the increase in the rotational correlation time. Our NMR data shows that the peptide adopts a helical conformation in solution and the combination of STD and trNOE data support that that this conformation is maintained while bound to the shell protein PduK.

Fluorescein-labeled P18 as a reporter

To visualize the interaction between the BMC shell and the P18 peptide we have used fluorescence microscopy. The P18 peptide does not contain any lysine residues and reaction with fluorescence isothiocyinate (FITC) produces a singularly labeled peptide at the N-terminal NH_3 group that is easily purified in high yield. Empty microcompartments, consisting of the shell proteins PduABJKNU, which were either unlabeled or contained PduA labeled with the fluorescent mCherry tag were purified and incubated with the FITC labeled P18 peptide. Following centrifugation to pellet the BMCs and remove any unbound P18 the BMCs were

resuspended in buffer and visualized using a widefield microscope and optisplit (Cairn Research LTD) with appropriate filters (Supplementary Figure 8). Localization of green fluorescence was observed with the purified microcompartments. When the mCherry labeled compartments were used, co-localization of the two fluorophores was observed, confirming that the FITC-P18 peptide is binding to the BMC shell. From these images it is not possible to determine if the peptide is binding to the outside or the inside of the BMCs. It is presumed that PduP must be located in the lumen and therefore unlikely to interact with the outside of the shell, however, it could be possible that the peptide is able to access the interior of the microcompartment through one of the pores formed by the shell proteins. Such pores are large enough to allow access to cofactors such as cobalamin, coenzyme A and NAD^+.

Internalization of a reporter protein using P18

To probe the internalization of proteins fused to the signaling sequence PduP18 into a BMC at higher resolution we used a reporter protein that can be visualized by electron microscopy. Here, we fused the modified pea ascorbate peroxidase (APEX) (32) to the C-terminus of the P18 peptide. APEX allows for high resolution imaging of subcellular structures containing the enzyme by its ability to catalyze the H_2O_2 dependent polymerization of diaminobenzidine, generating an insoluble precipitate that stains with OsO_4. Thin sections of the strain producing the P18-APEX fusion protein and the shell proteins PduABJKNU were generated and viewed under TEM to reveal electron dense areas that match the shape and size of microcompartments (Figure 4) indicating that APEX is active when fused to P18 and that it is located internally within the BMC where it catalyzes DAB polymer formation. No similar electron dense areas were observed in sections of a control strain that contained the genes for the empty BMC and APEX without the P18 targeting peptide. The result suggests that both H_2O_2 and DAB are able to enter the BMC, where the polymerized product precipitates.

Construction of an ethanol bioreactor

To generate the proof of principle that it is possible to incorporate pathways not natively associated with microcompartments into recombinant empty bacterial Pdu microcompartments we chose to target the enzymes for the transformation of pyruvate to ethanol to an empty BMC. Ethanol production inside microcompartments requires the heterologous expression of pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adh*) and their targeting to the inside of the protein shell (Figure 1). In this study the *pdc* and *adh* are derived from the strictly anaerobic ethanogenic Gram-negative bacterium *Z. mobilis*. The *pdc* gene was cloned so as to allow it to

encode for an N-terminal fusion with the P18 targeting sequence. The *adh* gene was also initially cloned so as to encode a fusion with the first 18 amino acids of PduD (a D18 targeting sequence). Although the first 18 amino acids of PduD are sufficient for targeting proteins (26) we also decided to explore an extended PduD sequence to allow for more space between PduD18 and the enzyme and thus chose to fuse the first 60 amino acids of PduD to Adh. The secondary structure prediction and SWISS-MODEL results of *C. freundii* PduD indicated that the first 18 residues are followed predominantly by a coiled secondary structure.

These constructs encoding the P18-Pdc D18-Adh, and P18-Pdc D60-Adh were investigated for their ability to target the enzymes efficiently to the BMC. The respective strains and their controls were cultured for 96 hours in LB medium supplemented with 4 % glucose. Growth rates and relative ethanol contents per OD_{600}=1 were recorded. Our growth studies showed that the co-expression of the shell protein construct (pLysS-*pduABJKNU*) and *pdc* and *adh* (tagged and untagged) caused a growth delay up to 48 hours compared to strains that only produced *pdc* and *adh* and no genes for shell proteins. At 48 hours however, strains producing shell proteins had reached a similar OD_{600} as their control strains that did not produce shell proteins (Supplementary Figure 9). The ethanol production strains without shell protein construct produced similar amounts of ethanol (Pdc Adh: 36 ± 5 mM ethanol, P18-Pdc D18-Adh: 44 ± 11 mM ethanol and P18-Pdc D60-Adh: 38 ± 1 mM ethanol). However, when shell proteins were co-produced the two strains containing Pdc and Adh with signalling sequences produced more ethanol than the strain containing Pdc and Adh without signalling sequences (P18-Pdc D18-Adh: 59 ± 4 mM ethanol and P18-Pdc D60-Adh: 46 ± 1 mM ethanol versus Pdc Adh: 38 ± 8 mM) (Figure 5, Supplementary Figure 10).

Increased ethanol production in strains producing enzymes with signalling sequences seemed to be dependent on the production of shell proteins. At 48 hours, when the shell proteins were most abundant (Supplementary Figure 10), strains producing shell proteins and either P18-Pdc D18-Adh or P18-Pdc D60-Adh increased sharply in ethanol content and produced more ethanol than the control strains producing shell proteins and Pdc Adh (Figure 5). SDS PAGE analysis also suggested that less P18-Pdc was present in strains that produced shell proteins compared to strains that did not produce shell proteins (Supplementary Figure 10). Thus, the enzymes (with and without signalling sequences) are functional and apparently similarly active in solution. However, when targeting Pdc and Adh to microcompartments ethanol production was improved suggesting the enzymes were more active when targeted to microcompartments than in solution.

Purification of recombinant bioreactors

The activity of BMCs containing the Pdc and Adh were investigated by purifying the organelles. Bacteria containing the recombinant BMCs were lysed using Yeast Protein Extraction Reagent (YPER Plus) and the metabolosomes were separated from contaminating proteins on the basis of their differential solubility as described under microcompartment purification in the methods section. The solubility of the BMC appears to be dependent on the NaCl concentration of the buffer (Figure 6). Low salt content (20 mM NaCl) increased their solubility, whereas higher salt content (80 mM NaCl) lowered their solubility. Analysis of the final protein fraction of purified empty microcompartments using TEM and AFM confirmed the presence of apparently intact irregularly shaped polygonal recombinant microcompartments around 100 nm in diameter (Figures 7a and 7b).

Although, not completely homogeneous, the purified microcompartment fraction was found to be enriched in shell proteins PduA, -B, -B′, -J, -K, -U when analyzed by SDS-PAGE (Figure 7c). PduN was not detectable in this profile, but as it is suspected to act as a vertex protein rather than making up the facets of the structure, it is likely to be present in only very small quantities.

When untagged Pdc and Adh are produced in the presence of an empty BMC a small band at 64 kDa is present in the protein profile of microcompartments from this strain (Figure 7c). This band corresponds to the theoretical mass of Pdc (no band corresponding to the theoretical mass of Adh is visible). This suggests that some Pdc is either incorporated into the BMC in the absence of a signaling sequence or that some denatured Pdc co-purifies with the metabolosome.

Purified microcompartments from strains producing Adh and Pdc with signaling sequences (P18-Pdc D18-Adh and P18-Pdc D60-Adh) co-purified with Pdc and Adh proteins (Figure 7c) indicating that both Pdc and Adh were interacting with the BMC through their signaling sequences. To determine if the purification protocol was specific for microcompartments, untagged and tagged Adh and Pdc were purified in the absence of the shell proteins PduA, -B, -B′, -J, -K, -U using the described purification method. No Adh or Pdc protein bands were detected in the final fraction of the purification (Supplementary Figure 11) suggesting that the Adh and Pdc can only be purified with microcompartments.

A functional ethanol bioreactor

Purified microcompartments from the various strains were analyzed for their specific activity in the oxidation of NADH using a coupled assay that required both Pdc and Adh to be present and active (Figure 7d). Microcompartments from the strain containing the P18-Pdc and D60-Adh had the highest NADH oxidation activity of 215 nmol NADH oxidized min^{-1} mg^{-1}, which was approximately 20 times higher than the specific activity of microcompartments extracted from cells producing untagged Adh and Pdc. Microcompartments purified from the strain producing the P18-Pdc and D60-Adh were twice as active as microcompartments from the strain producing P18-Pdc and D18-Adh. It is possible that the shorter PduD18 sequence might restrain Adh and thus lead to lower activity *in vitro*. To confirm that ethanol is produced by the BMCs they were incubated overnight with pyruvate, NADH, thiamine pyrophosphate and $MgCl_2$ and the amount of ethanol was determined by GC-MS. Similar ethanol levels were detected in incubations containing microcompartments from the P18-Pdc D60-Adh strain (34 ± 3.4 mM) and the P18-Pdc D18-Adh strain (36 ± 0.2 mM), corresponding to turnovers of 68 % and 72 % respectively. Microcompartments from the strain containing untagged Pdc and Adh produced also some ethanol (23 ± 1.1 mM ethanol, 46 % turnover), which was probably due to some contaminating protein present in the purified sample. Microcompartments purified from the control strain without Pdc and Adh did not produce detectable levels of ethanol.

Discussion

The structure of the N-terminal region of PduP, P18, has been determined by NMR, revealing a strong helical structure. This structure is preformed and does not involve any induced fit with its binding partner (PduK) indicating a more general rather than a specific form of interaction. The fact that no structural change is required for binding suggests that the peptide adopts this conformation to reduce the entropic barrier for the interaction. Saturation transfer difference NMR has highlighted that one face of the helix shows the greatest interaction, involving several conserved hydrophobic residues. In particular the LIRNIL region contains the major molecular recognition features that allow for interaction with PduK. Our research has not identified the region of PduK to which the peptide binds. However, through analysis of C-terminal truncations we know that the binding region must lie in the N-terminal BMC domain between amino acids 1-96. Predicative modeling has suggested the C-terminal helix in PduA from *S. enterica* is responsible for the interaction with the internalized protein, PduP (30). Similarly, the truncated PduK (PduK96) protein contains a homologous helix in this region that could be the potential site of interaction. However, molecular detail on the interaction between the shell and the proteins which are encapsulated will require high resolution structural data.

We have demonstrated that PduP is able to interact with the shell protein PduK and shown that P18 can bind to PduK with an affinity in the sub micromolar range, consistent with tight binding. This level of binding explains how the proteins can interact but does not offer any insight into what happens to the targeted protein. Using isolated empty BMCs, which appear intact as viewed by TEM and AFM, addition of fluorescein-labeled P18 peptide results in localization of the peptide to the compartments *in vitro*. This likely reflects the strong interaction between P18 and the shell of the recombinant microcompartment. Evidence that the P18 targeting sequence leads to sequestration of the protein to the inside of the microcompartment comes from *in vivo* analysis of cells producing P18-tagged to an engineered ascorbate peroxidase (APEX) (32). The APEX enzyme is able to polymerize diaminobenzidine to give an insoluble polymer that stains with OsO_4 and therefore regions of the cells where APEX is present appear electron dense during imaging. Using this approach we have been able to show that dark regions of cell appear that are in agreement with the presence of BMCs, and consistent with the targeting and internalization of P18-tagged-APEX to the recombinant BMC. If the APEX was only associated with the outside of the BMC then we would have expected to see a darker staining around the periphery of the compartment.

It is not known how internalization of the P18-tagged protein occurs. One possibility, similar to the model suggested for carboxysome construction (4), is that a lattice of shell and P18-tagged protein forms simultaneously through interactions between P18 and PduK during assembly. It could also be that the targeting sequences encourage oligomerization around which the BMC forms. Alternatively, the targeting sequence could interact with a preformed BMC resulting in the inversion of a PduK-containing tile that results in the import of the tag-containing protein to the inside of the compartment. However, our data do not discriminate between these possibilities.

We sought to see if the P18 targeting sequence could be used to direct a small pathway to an empty BMC as proof of concept that new pathways could be engineered into microcompartments. We sought to target a pyruvate decarboxylase and alcohol dehydrogenase to the BMC since this would allow for ethanol production. We found that the P18 sequence allowed for the targeting of the *Z. mobilis* Pdc to the BMC but strangely the P18 did not work well with the *Z. mobilis* Adh, as it was not possible to obtain a stable *E. coli* strain containing the shell protein construct and a P18-*pdc* P18-*adh* construct. We thus chose another targeting sequence, the N-terminal region of PduD and constructed two different tags, one of 18 amino acids (D18) and a longer one of 60 amino acids (D60). Both these sequences allowed for the targeting of Adh to the BMC. Although microcompartments isolated from the strain producing the longer sequence were more active *in vitro* they produced similar amount of ethanol as the BMCs containing the shorter D18-Adh. *In vivo*, the strain containing D60-Adh produced less ethanol than the D18-Adh strain. It is possible that the two signalling sequences lead to different levels of incorporation into microcompartments over time. By combining various constructs we were able to show that it is possible to direct both P18-Pdc and D18-Adh or D60-Adh to the same BMC and produce a microcompartment that has the ability to convert pyruvate into ethanol. Isolated microcompartments containing Pdc and Adh must be able to take up pyruvate and thiamine to be metabolically active, suggesting that there is a lack in specificity in the pores of the compartment. The compartment, does increase the ethanogenic potential for *E. coli* as the amount of ethanol produced by cells containing compartmentalized enzymes is higher than for cells expressing cytoplasmically-located enzymes. ~~However, this is perhaps not surprising as the yield limitation in ethanogenic *E. coli* is thought to be Pdc and Adh enzyme toxicity to ethanol, so on this basis compartmentalization is unlikely to enhance ethanol production (33).~~

In conclusion, we have determined the structure of a BMC-targeting peptide (P18) and identified its major recognition epitopes. We have shown how this peptide from PduP and a related peptide from PduD are able to target enzymes to a recombinant BMC. We have used a high-

resolution ascorbate peroxidase (APEX) system to demonstrate that the P18 peptide results in the internalization of the protein cargo within the BMC and by using these targeting sequences we have generated a recombinant ethanol-producing bioreactor. This report therefore provides evidence that BMCs can be easily manipulated for the construction of new purposes and hold significant potential especially for processes that involve toxic intermediates. BMCs represent a subcellular compartment that is apposite for the redesign of biological processes.

Materials and Methods

Bacterial strains and media

Strains and plasmids used in this study are listed in Supplementary Table 4. *E. coli* JM109 was used for routine cloning procedures. *E. coli* strain BL21(DE3) was used for the expression of the genes for ethanol production (*pdc* and *adhB from Z. mobilis*) and the genes encoding the shell proteins PduA, B, B', J, K, N, U. Strains were grown in Luria–Bertani (LB) medium supplemented with antibiotics (50 µg/ml ampicillin, 35 µg/ml chloramphenicol, chloramphenicol stock dissolved in methanol at 10 mg/ml) as needed. All strains were grown at 37 °C and 160 r.p.m. Isolated colonies of *E. coli* strains were grown overnight in liquid medium and used as inoculum into fresh medium at a starting OD_{600}=0.05, unless otherwise indicated. At OD_{600}=0.8 protein production was induced with 400 µM IPTG overnight at 18 ºC. Growth rates were recorded in a volume of 1 ml in 24 well plates (Greiner) using a BMG Labtech FLUOstar OPTIMA plate reader in absorbance mode at 600 nm for 40 hours.

Pull down assays

Pull down assays were conducted between PduP and the seven individual shell proteins. Two constructs were co-expressed in *E. coli* BL21(DE3). The first plasmid (pET14b) contained one of the shell proteins as an N-terminal his-tag fusion. The second plasmid (pLysS) contained the gene encoding PduP without any tag. The transformants were grown in 1L of Luria-Bertani medium to an OD_{600} of 0.8 when protein production was induced with 400 mM IPTG at 16 ºC overnight. The cells were harvested by centrifugation at 4,000 x g at 4 ºC for 10 minutes and resuspended in 10 ml of binding buffer (20 mM Tris-HCl pH 8.0, 0.5 M NaCl, 10 mM imidazole). The cells were lysed by sonication (six 30 seconds bursts, with 30 second cooling intervals on ice at an amplitude of 65 %) and insoluble debris removed by centrifugation at 37, 000 x g for 15 minutes. The supernatant was applied to a chelating sepharose column charged with Ni^{2+}. Once bound to the column, the crude extract was subjected to washes containing stepwise increases in imidazole concentration up to 400 mM. All wash fractions were analyzed by SDS-PAGE to confirm co-elution of the His-tagged and untagged proteins. The identities of the eluted proteins were confirmed by peptide mass fingerprinting. As a negative control, the His-tagged shell protein was omitted and under these conditions no PduP was present in the elution fractions (as determined by SDS-PAGE).

NMR

Peptide structural NMR datasets were obtained at 283 K using a 14.1 T (600 MHz ^{1}H) Bruker Avance III NMR spectrometer equipped with a QCI-F cryoprobe. All NMR samples were 330 µL within a Shigemi NMR tube and contained 1 mM P18 peptide in 20 mM potassium phosphate buffer at pH 7.0 containing 100 mM sodium chloride and 9% (v/v) deuterium oxide. NMR data processing was completed using TopSpin 3.1 (Bruker), assignments were completed using CCPN Analysis (34). Data sets for structural determination were collected in the absence and presence of 30 % (v/v) trifluroethanol-d_3 (TFE). ^{1}H chemical shifts and through-space structural assignments were obtained from two-dimensional TOCSY and NOESY NMR experiments with mixing times of 80 ms and 350 ms respectively (Supplementary Figure 3; Supplementary Table 1). Structural ensembles were calculated using CNS (35) and included dihedral angles confirmed using DANGLE (36). The final ensemble of each structure was water-minimized using YASARA Structure software (available from http://www.yasara.org). Supplementary Table 2 refinement statistics together with Supplementary Figures 6 and 7 confirm the precision and quality of each peptide structure.

Standard ^{1}H STD NMR experiments with a 3-9-19 WATERGATE were performed using 1 mM P18 peptide with 15 µM PduK in the same PBS buffer system used for structure calculations in the absence of TFE. Gaussian STD excitation pulses of 20 ms duration and a γB_1 of 140 Hz, were applied for 2 s at -3 ppm and -30 ppm for saturation and control respectively with on/off interleaved and extracted after the experiment (Supplementary Figure 5). The experiment, when repeated with saturation at +13 ppm and control at -30 ppm, produced the same overall result but at lower intensity due to reduced spin diffusion. ^{1}H T_1 relaxation time constants for the P18 peptide were obtained using an inversion-recovery sequence that included a 3-9-19 WATERGATE. ^{1}H T_1 values were then used to obtain quantitative STD NMR (qSTD) amplification factors using the GEM-CRL method described previously (37). Intensities for STD and T_1 relaxation experiments obtained using Spectrus Processor (ACD Labs)

2D trNOE spectra were measured at mixing times of 50, 100, 150, 200, 250, 300, 350 and 400 ms using 1 mM P18 peptide and 66 µM PduK (15-fold ligand excess) in 20 mM potassium phosphate buffer at pH 7.0 containing 100 mM sodium chloride and 9% (v/v) deuterium oxide.

Internalization of APEX using P18

To probe internalization of a protein fused to PduP18 into recombinant microcompartments the electron microscopy (EM) reporter protein APEX (32) was fused to the C-terminus of PduP18 (pET14b-*pduP18-APEX*). The control plasmid contained APEX alone (pET14b-*APEX*). APEX is

a monomeric and activity enhanced mutant of pea ascorbate peroxidase and is expressed and reacted in live cells where it catalyzes the oxidative polymerization of diaminobenzidine (DAB) to generate a cross-linked and locally deposited precipitate. EM contrast is created when the DAB polymer is stained with electron dense OsO_4. The protocol was adapted from Martell et al as follows. The *E.coli* strain BL21(DE3) was co-transformed with pLysS-*pduABB'JKNU*, the plasmid encoding the shell proteins that form empty microcompartments and pET14b-*APEX* or pET14b-*pduP18-APEX*. 10 ml LB starter cultures containing appropriate antibiotics and 1 colony were cultured for 6 hours at 37 ºC at 160 rpm and then inoculated into 50 ml 2YT at starting OD_{600}=0.01. The strains were incubated for 16 hours at 37 ºC at 160 rpm and then harvested by centrifugation. 0.05 g of wet cells were washed 3 x 5 min in 1 ml of 100 mM sodium cacodylate, 2 mM $CaCl_2$, pH 7.4. The cells were re-suspended in 1 ml of freshly made solution of 0.5 mg/ml (1.4 mM) DAB tetrahydrochloride, combined with 0.03 % (v/v) H_2O_2 in chilled 100 mM sodium cacodylate, 2 mM $CaCl_2$, pH 7.4. After 1 min incubation the cells were washed 3 x 5 min in 100 mM sodium cacodylate, 2 mM $CaCl_2$, pH 7.4 followed by 60 min fixation at room temperature in 1 ml of 2% glutaraldehyde in 100 mM sodium cacodylate, 2 mM $CaCl_2$, pH 7.4. The cells were washed 3 x 5 min in 1 ml chilled buffer, followed by post fixation staining with 2% osmium tetroxide for 60 min in chilled buffer. The cells were rinsed 3 x 5 min in chilled distilled water and placed in 30 % ethanol overnight at 4 ºC. The samples were dehydrated in a cold graded ethanol series, 10 min each (50%, 70%, 90%, 100%, 100%) and then washed once in room temperature anhydrous ethanol (to avoid condensation) followed by infiltration with LV resin (Agar Scientific): 1:1 (v/v) anhydrous ethanol and resin for 60 min. The cells were incubated twice in 100% LV resin for 1 hour and finally transferred into fresh resin and polymerized at 60ºC for 16 hours. 85 nm sections were collected on 300 mesh copper grids and visualized under the TEM as described previously (24).

Construction of plasmids for ethanol production

Pdc was amplified by PCR using 5'-CATCATATGAGTTATACTGTCGG-3' (forward primer) and 5'-CATGAATTCAAAACTAGTCAGAGGAGCTTGTTAACAGGC-3' (reverse primer) introducing restriction sites *Nde*I (5') and *Spe*I and *EcoR*I (3') (all underlined). *Pdc* was cloned via *Nde*I and *EcoR*I sites into a modified pET23b-P18 (pSF59) allowing for the fusion of P18 and the 5' end of *pdc*, creating pSN5. *Adh* was amplified by PCR using 5'-CACGAGCTCATGGCTTCTTCAACTTTTTATATTCC-3' (forward primer) and 5'-CACGAATTCAAAACTAGTCAGAAAGCGCTCAGGAAGAGTTC-3' (reverse primer) introducing restriction sites *Sac*I (5') and *Spe*I and *EcoR*I (3'). *Adh* was cloned via *Sac*I and *Spe*I sites into

pET3a containing *D18* and *D60* signaling sequences allowing for the fusion of the 3' end of *D18* or *D60* and the 5' end of *adh*, creating pSF61 and pSF63, respectively. Finally, two constructs were built housing *D18-adh* and *P18-pdc* (pSF64) and *D60-adh* and *P18-pdc* (pSF65). A plasmid with *adh* and *pdc* (pSN4) only was constructed as negative control.

Solubility study of the shell proteins

BL21(DE3) was transformed with pLysS-*mCherry-pduA-BB'JKNU*, the plasmid encoding the shell proteins that form empty microcompartments (24) with PduA fused to a red fluorescent mCherry tag. The strain was cultured and the cells lysed as described under section 'Microcompartment purification'. The cell lysate was aliquoted into eight 2 ml samples and centrifuged (5 min at 11,300 x g, 4 °C). The pellets were re-suspended in 1 ml 20 mM Tris-HCl, pH= 8.0 containing increasing concentrations of NaCl (0, 20, 40, 60, 80, 100, 200, 500 mM). The suspension was pelleted (5 min at 11,300 x g, 4 °C) and supernatants and pellets were collected and analyzed by SDS-PAGE. The mCherry tag allowed the visualization of the shell proteins.

Microcompartment purification

BL21(DE3) strains containing genes for BMC formation were cultured in 200 ml LB medium and the cells were harvested by centrifugation for 10 minutes at 4 °C at 2683 x g. 1 g wet cell pellet was re-suspended in 10 ml YPER Plus (Pierce) supplemented with 1 tablet of Complete Protease Inhibitor Cocktail (Roche) and 1250 units Benzonase® Nuclease (YPB) and incubated for 3 hours at room temperature under gentle agitation.

As outlined above we observed that the shell proteins varied in solubility depending on the salt concentration of the buffer they were re-suspended in. This property was used to separate the shell proteins from other proteins by subjecting the samples to varying salt concentrations as follows: The lysate was pelleted for 5 min at 11,300 x g (Beckman Coulter, rotor JA 25.50) and the microcompartment containing pellet was collected. The pellet was re-suspended in 2 ml 20 mM Tris-HCl, pH 8.0, 20 mM NaCl. The suspension was centrifuged at 4 °C for 5 min at 11,000 x g (Multispeed Refrigerated Centrifuge, ALC PK121R) and the supernatant which contained the solubilized microcompartments was collected. The NaCl concentration of the supernatant was raised to 80 mM by addition of 5 M NaCl. The microcompartments became insoluble and the cloudy solution was centrifuged at 4 °C for 5 min at 11,000 x g. The pellet with the microcompartments was collected, re-suspended in 1 ml of 20 mM Tris-HCl, pH 8.0 and finally

clarified by centrifugation (5 min at 11,000 x g). The final supernatant was collected for further analysis.

EM and AFM

For EM analysis 10 µl purified microcompartments were mounted onto formvar, carbon coated 600 mesh copper grids for 2 minutes, followed by the addition of 20 µl 2.5 % (v/v) glutaraldehyde in PBS for 1 minute. The grids were washed three times in 20 µl drops of 2.5 % (v/v) glutaraldehyde in PBS, and then three times in water. Finally the grids were dried and stained with 2 % aqueous uranyl acetate.

AFM images were collected on a Bruker Multimode 8 scanning probe microscope. Purified BMCs were deposited on a graphite substrate by incubating 20 µl of sample on a freshly cleaved HOPG surface for 5 min. The sample was then fixed using 2.5 % (v/v) glutaraldehyde in PBS, and dried with a gentle stream of N_2 gas. Images were collected in air using the peak-force tapping mode with peak-asyst air cantilever probe (Bruker) with a nominal spring constant of 0.4 N/m, and processed using the supplied Nanoscope software.

Activity Assays

The activities of Pdc and Adh were monitored spectrophotometrically by measuring the absorbance change at 340 nm due to the oxidation of NADH. The assay depended on the conversion of pyruvate to acetaldehyde, catalyzed by Pdc and required Adh as the coupling enzyme to facilitate NADH oxidation. The reaction was carried out at room temperature in 20 mM Tris-HCl, pH 8.0. 1 ml reactions contained 5 mM pyruvate, 0.15 mM NADH, 5 mM $MgCl_2$, 0.1 mM thiamine pyrophosphate and 481 µg purified microcompartments. The reaction was started by addition of pyruvate and NADH and the rate of NADH oxidation was measured at 340 nm.

Analysis of ethanol production

In vivo ethanol production was determined by analysis of the growth medium after culturing the strains. 500 ml LB cultures supplemented with 4 % glucose were cultured for 96 hours at 28 ºC, shaking at 160 rpm. 2 ml samples were taken for OD_{600} readings and GC analysis at 1, 2, 3, 4, 5, 6, 10, 24, 48, 72 and 96 hours. Cells were removed by centrifugation and the growth medium subjected to GC-MS analysis.

Purified microcompartments were tested for their ability to convert pyruvate to ethanol by overnight incubation at room temperature in 1 ml reactions containing 50 mM pyruvate, 50 mM NADH, 50 mM $MgCl_2$ and 1 mM thiamine pyrophosphate and 1.6 mg protein. The reactions were clarified by centrifugation for 10 minutes and the supernatant was filtered through a vivaspin column (molecular cut off 10,000 Da). The filtrate was analyzed for ethanol content by GC-MS.

As commercial NADH (Sigma) contains traces of ethanol, NADH used in the reaction was produced using alternative published methods (38-40) by incubation of 0.03 M glucose, 0.015 M NAD^+ and 35 U Glucose dehydrogenase in 0.05 M sodium bicarbonate, pH 7.5 at 37 ºC overnight. NADH was purified on DEAE in a 0-0.2 M ammonium bicarbonate gradient. After two cycles of rotary evaporation NADH was re-suspended in Millipore water.

Acknowledgements

This work was supported by grants from the Biotechnology and Biological Sciences Research Council (BBSRC) BB/H013180/1 and BB/E024203/1 and The Welcome Trust 091163/Z/10/Z. We thank K. Howland for synthesis of the P18 peptide and Luke Lloyd for cloning of the APEX gene.

References

1. Frank, S., Lawrence, A. D., Prentice, M. B., and Warren, M. J. (2013) Bacterial microcompartments moving into a synthetic biological world, *J Biotechnol 163*, 273-279.
2. Cannon, G. C., Bradburne, C. E., Aldrich, H. C., Baker, S. H., Heinhorst, S., and Shively, J. M. (2001) Microcompartments in prokaryotes: carboxysomes and related polyhedra, *Applied and environmental microbiology 67*, 5351-5361.
3. Iancu, C. V., Ding, H. J., Morris, D. M., Dias, D. P., Gonzales, A. D., Martino, A., and Jensen, G. J. (2007) The structure of isolated Synechococcus strain WH8102 carboxysomes as revealed by electron cryotomography, *Journal of molecular biology 372*, 764-773.
4. Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C., and Jensen, G. J. (2010) Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells, *Journal of molecular biology 396*, 105-117.
5. Kerfeld, C. A., Heinhorst, S., and Cannon, G. C. (2010) Bacterial microcompartments, *Annual review of microbiology 64*, 391-408.
6. Schmid, M. F., Paredes, A. M., Khant, H. A., Soyer, F., Aldrich, H. C., Chiu, W., and Shively, J. M. (2006) Structure of Halothiobacillus neapolitanus carboxysomes by cryo-electron tomography, *Journal of molecular biology 364*, 526-535.
7. Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C., and Yeates, T. O. (2008) Atomic-level models of the bacterial carboxysome shell, *Science 319*, 1083-1086.
8. Havemann, G. D., and Bobik, T. A. (2003) Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar Typhimurium LT2, *J Bacteriol 185*, 5086-5095.
9. Havemann, G. D., Sampson, E. M., and Bobik, T. A. (2002) PduA is a shell protein of polyhedral organelles involved in coenzyme B-12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar typhimurium LT2, *J Bacteriol 184*, 1253-1261.
10. Kofoid, E., Rappleye, C., Stojiljkovic, I., and Roth, J. (1999) The 17-gene ethanolamine (eut) operon of Salmonella typhimurium encodes five homologues of carboxysome shell proteins, *J Bacteriol 181*, 5317-5329.
11. Parsons, J. B., Dinesh, S. D., Deery, E., Leech, H. K., Brindley, A. A., Heldt, D., Frank, S., Smales, C. M., Lunsdorf, H., Rambach, A., Gass, M. H., Bleloch, A., McClean, K. J., Munro, A. W., Rigby, S. E., Warren, M. J., and Prentice, M. B. (2008) Biochemical and structural insights into bacterial organelle form and biogenesis, *J Biol Chem 283*, 14366-14375.
12. Stojiljkovic, I., Baumler, A. J., and Heffron, F. (1995) Ethanolamine utilization in Salmonella typhimurium: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster, *J Bacteriol 177*, 1357-1366.
13. Brinsmade, S. R., Paldon, T., and Escalante-Semerena, J. C. (2005) Minimal functions and physiological conditions required for growth of salmonella enterica on ethanolamine in the absence of the metabolosome, *Journal of bacteriology 187*, 8039-8046.
14. Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S., and Aldrich, H. C. (1999) The propanediol utilization (pdu) operon of Salmonella enterica serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-propanediol degradation, *J Bacteriol 181*, 5967-5975.
15. Sriramulu, D. D., Liang, M., Hernandez-Romero, D., Raux-Deery, E., Lunsdorf, H., Parsons, J. B., Warren, M. J., and Prentice, M. B. (2008) Lactobacillus reuteri DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1,2-propanediol by disproportionation, *J Bacteriol 190*, 4559-4567.

16. Crowley, C. S., Cascio, D., Sawaya, M. R., Kopstein, J. S., Bobik, T. A., and Yeates, T. O. (2010) Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell, *J Biol Chem 285*, 37838-37846.
17. Yeates, T. O., Crowley, C. S., and Tanaka, S. (2010) Bacterial microcompartment organelles: protein shell structure and evolution, *Annual review of biophysics 39*, 185-205.
18. Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C., and Shively, J. M. (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments, *Nature reviews. Microbiology 6*, 681-691.
19. Yeates, T. O., Thompson, M. C., and Bobik, T. A. (2011) The protein shells of bacterial microcompartment organelles, *Current opinion in structural biology 21*, 223-231.
20. Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D., and Yeates, T. O. (2013) Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins, *Protein science : a publication of the Protein Society 22*, 660-665.
21. Kerfeld, C. A., Sawaya, M. R., Tanaka, S., Nguyen, C. V., Phillips, M., Beeby, M., and Yeates, T. O. (2005) Protein structures forming the shell of primitive bacterial organelles, *Science 309*, 936-938.
22. Tanaka, S., Sawaya, M. R., and Yeates, T. O. (2010) Structure and mechanisms of a protein-based organelle in Escherichia coli, *Science 327*, 81-84.
23. Pang, A., Liang, M., Prentice, M. B., and Pickersgill, R. W. (2012) Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB, *Acta Crystallogr D Biol Crystallogr 68*, 1642-1652.
24. Parsons, J. B., Frank, S., Bhella, D., Liang, M., Prentice, M. B., Mulvihill, D. P., and Warren, M. J. (2010) Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement, *Mol Cell 38*, 305-315.
25. Bonacci, W., Teng, P. K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P. A., and Savage, D. F. (2012) Modularity of a carbon-fixing protein organelle, *Proc Natl Acad Sci U S A 109*, 478-483.
26. Fan, C., and Bobik, T. A. (2011) The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment, *J Bacteriol 193*, 5623-5628.
27. Fan, C. G., Cheng, S., Liu, Y., Escobar, C. M., Crowley, C. S., Jefferson, R. E., Yeates, T. O., and Bobik, T. A. (2010) Short N-terminal sequences package proteins into bacterial microcompartments, *Proc Natl Acad Sci U S A 107*, 7509-7514.
28. Kinney, J. N., Salmeen, A., Cai, F., and Kerfeld, C. A. (2012) Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly, *The Journal of biological chemistry 287*, 17729-17736.
29. Meyer, B., and Peters, T. (2003) NMR Spectroscopy techniques for screening and identifying ligand binding to protein receptors, *Angewandte Chemie-International Edition 42*, 864-890.
30. Fan, C., Cheng, S., Sinha, S., and Bobik, T. A. (2012) Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments, *Proc Natl Acad Sci U S A 109*, 14995-15000.
31. Povey, J. F., Smales, C. M., Hassard, S. J., and Howard, M. J. (2007) Comparison of the effects of 2,2,2-trifluoroethanol on peptide and protein structure and function, *Journal of Structural Biology 157*, 329-338.
32. Martell, J. D., Deerinck, T. J., Sancak, Y., Poulos, T. L., Mootha, V. K., Sosinsky, G. E., Ellisman, M. H., and Ting, A. Y. (2012) Engineered ascorbate peroxidase as a genetically encoded reporter for electron microscopy, *Nat Biotechnol 30*, 1143-1148.

33. Trinh, C. T., Huffer, S., Clark, M. E., Blanch, H. W., and Clark, D. S. (2010) Elucidating mechanisms of solvent toxicity in ethanologenic Escherichia coli, *Biotechnol Bioeng 106*, 721-730.
34. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins 59*, 687-696.
35. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr D Biol Crystallogr 54 (Pt 5)*, 905-921.
36. Cheung, M. S., Maguire, M. L., Stevens, T. J., and Broadhurst, R. W. (2010) DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure, *J Magn Reson 202*, 223-233.
37. Kemper, S., Patel, M. K., Errey, J. C., Davis, B. G., Jones, J. A., and Claridge, T. D. (2010) Group epitope mapping considering relaxation of the ligand (GEM-CRL): including longitudinal relaxation rates in the analysis of saturation transfer difference (STD) experiments, *J Magn Reson 203*, 1-10.
38. Gassner, G., Wang, L., Batie, C., and Ballou, D. P. (1994) Reaction of phthalate dioxygenase reductase with NADH and NAD: kinetic and spectral characterization of intermediates, *Biochemistry 33*, 12184-12193.
39. Hauck, T., Landmann, C., Raab, T., Bruhlmann, F., and Schwab, W. (2002) Chemical formation of 4-hydroxy-2, 5-dimethyl-3 [2H]-furanone from d-fructose 1, 6-diphosphate, *Carbohydrate research 337*, 1185-1191.
40. Margolis, S., Howell, B. F., and Schaffer, R. (1976) Purification and analysis of the purity of NADH, *Clinical chemistry 22*, 1322-1329.

Figure Legends

Figure 1. Steps required for the conversion of the 1,2-propanediol utilization (Pdu) microcompartment into an ethanol bioreactor. The Pdu microcompartment shell is built from hexameric tiles composed of PduA, B, J, K, U and T (purple) that form the facets of the structure whereas pentameric tiles (PduN, cyan) form the vertices. 1,2-Propanediol enters the shell through pores in the shell proteins and is metabolized to propionyl-CoA (red box), which leaves the compartment and is further converted to propionate. Enzymes that are encapsulated within the metabolosome contain short signaling peptides. Changing the specificity of the Pdu microcompartment is achieved by stripping out the Pdu pathway and replacing it with the required pathway *e.g*. ethanol production (green box). Fusion of signaling peptides to the new pathway enzymes facilitates internalization of the heterologous proteins. Abbreviations used in figure: 1,2-PD = 1,2-propanediol, PA = propionaldehyde, P-OH = 1-propanol, PCoA = propionyl-CoA, POI = protein of interest, SP = signalling peptide.

Figure 2. Solution structure of the P18 peptide and illustration of its interaction surface. (a) Solution structure of the P18 peptide highlighting the residues with the highest qSTD factors for the interaction between the peptide and the shell protein PduK. (b) Structure of the P18 peptide showing the solvent accessible surface colored as a percentage of the maximum STD transfer to highlight significant interactions across the face of the helix. (c) Tabular representation of the relative qSTD amplification factors for the P18/PduK interaction.

Figure 3. Transferred NOE build-up rates for the P18 peptide in the presence of PduK. Various NOE build-up curves for the P18 peptide in the absence (empty triangle) and presence (filled triangle) of PduK. (a) 4SerH to 3ThrH, (b) 6LeuH to 3ThrHg2*,(c) 6LeuH to 3ThrHa, (d) 8ThrH to 5GluHa, (e) 12AsnH to 13IleH, (f) 16GluHga to 13IleHg2*. The peptide to protein ratio is 15:1.

Figure 4. Internalization of an EM reporter protein using P18. Transmission electron microscopy images of thin sections of *E. coli* BL21(DE3) carrying the genes for an empty microcompartment (*pduABJKNU*) and the EM reporter protein APEX fused to P18 (top), APEX (middle) or P18-Pdc (bottom). Before embedding live cells were treated with DAB, the substrate for APEX, followed by staining with OsO_4. Sections of the P18-APEX strain show patches of electron density that match the size and shape of microcompartments indicating that APEX is internalized (right). These are not present in the APEX and P18-Pdc expressing strains. Scale bar is 0.5 µm.

Figure 5. *In vivo* ethanol production. The graph shows the ethanol content (mM) of the growth medium of *E. coli* strains producing shell proteins only (circle), shell proteins and untagged Pdc and Adh (triangles), shell proteins and tagged Pdc and Adh (P18-Pdc D18-Adh is represented by squares, P18-Pdc and D60-Adh is represented by diamonds). Ethanol concentration is normalized to OD_{600}=1 and plotted from 6 to 96 hours. Ethanol content was determined in triplicates.

Figure 6. Solubility of microcompartments with salt. To follow the microcompartments colorometrically PduA was tagged with mCherry and compartments composed of mCherry-PduABB'JKNU were produced in BL21(DE3). The effect of NaCl concentration on the solubility of these colored isolated empty micrompartments was followed. After cell lysis in YPER Plus the lysate was centrifuged and the resulting pellet was resuspended in 20mM Tris-HCl (pH=8.0) containing 0 to 500 mM NaCl. The suspension was centrifuged again resulting in supernatant and pellet fractions, which were then analysed by SDS-PAGE. Between 0 - 40 mM and at 500 mM NaCl the microcompartment proteins are soluble and are found in the supernatant fraction.

Between 60-200 mM NaCl the microcompartments are less soluble and are found in the pellet fraction.

Figure 7. Isolation of recombinant bioreactors and determination of activity. (a) TEM and (b) AFM (peak force error) images of isolated empty microcompartments (PduABB'JKNU) showing how they purify as intact units (scale bar 100 nm). (c) SDS page loaded with 10 µg microcompartments purified from four strains co-producing PduA, B, B', J, K, N, U and either no enzymes (shell only) or Pdc and Adh without signalling sequences (Pdc Adh) or P18-Pdc and D18-Adh or P18P-Pdc and D60-Adh. M is SDS7 marker. (d) Specific activity of Adh in nmol NADH oxidized min^{-1} mg^{-1}. Activities were determined in triplicates with purified microcompartments isolated from the four strains described in panel c.

Figure 1.

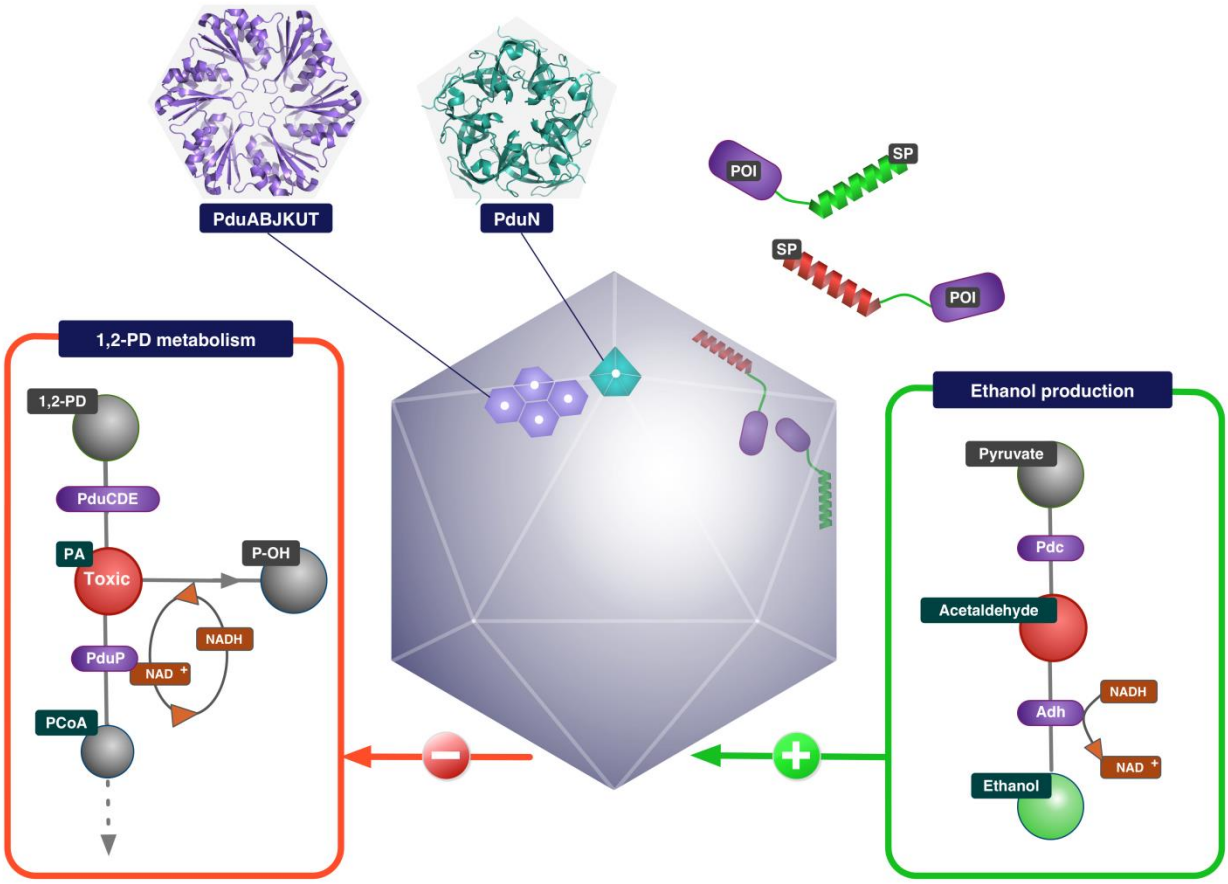

PduABJKUT
PduN
POI
SP
SP
POI
1,2-PD metabolism
1,2-PD
PduCDE
PA
Toxic
P-OH
NADH
PduP
NAD +
PCoA
Ethanol production
Pyruvate
Pdc
Acetaldehyde
NADH
Adh
NAD +
Ethanol

Figure 2.

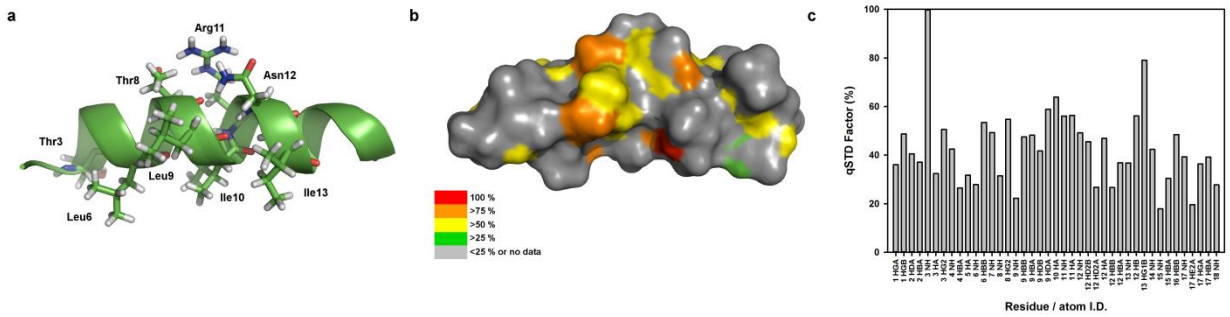

a
Arg11
Thr8
Asn12
Thr3
Leu9
Ile10
Ile13
Leu6
b
100 %
>75 %
>50 %
>25 %
<25 % or no data
c
qSTD Factor (%)
Residue / atom I.D.

Figure 3.

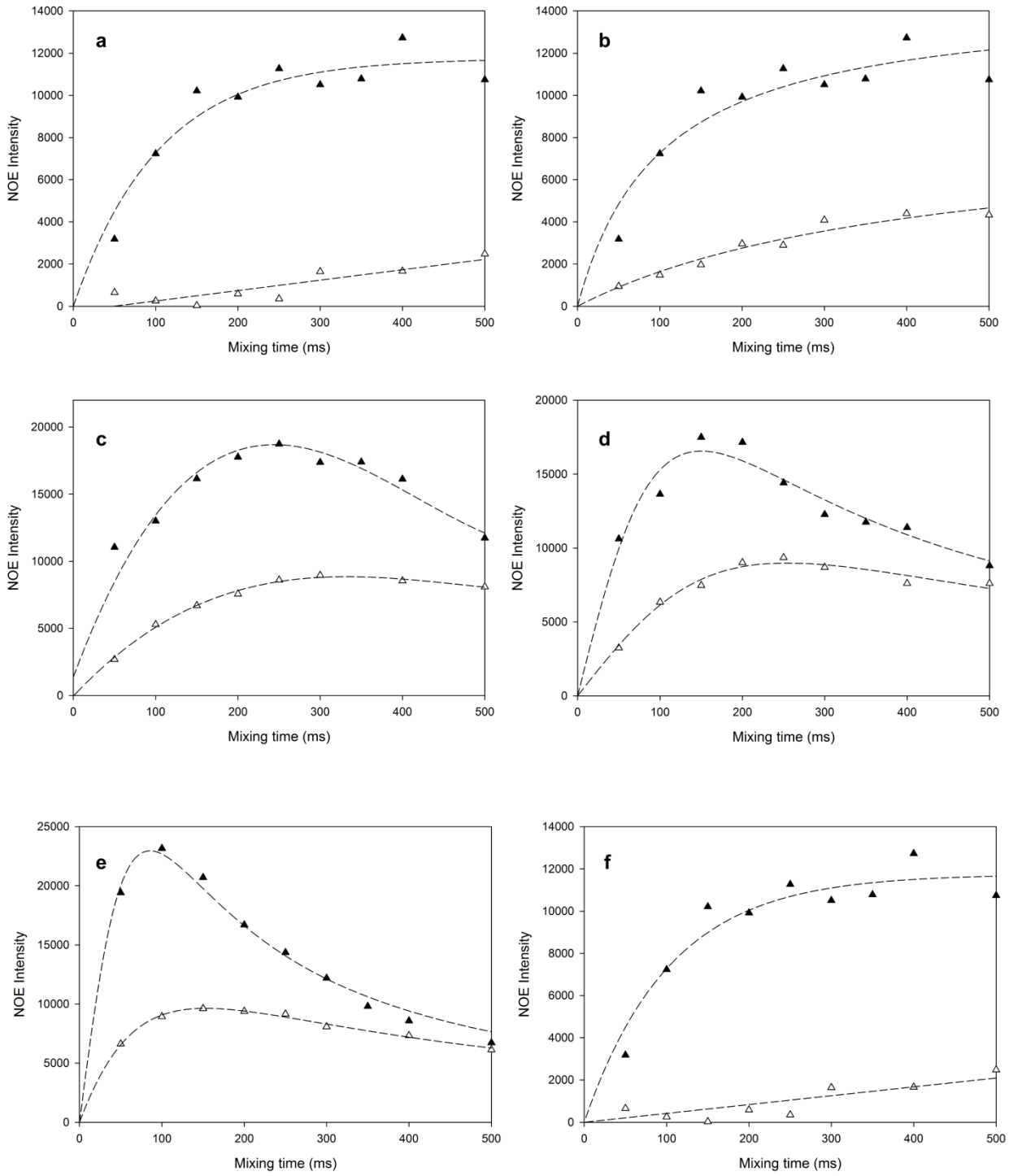


Figure 4.

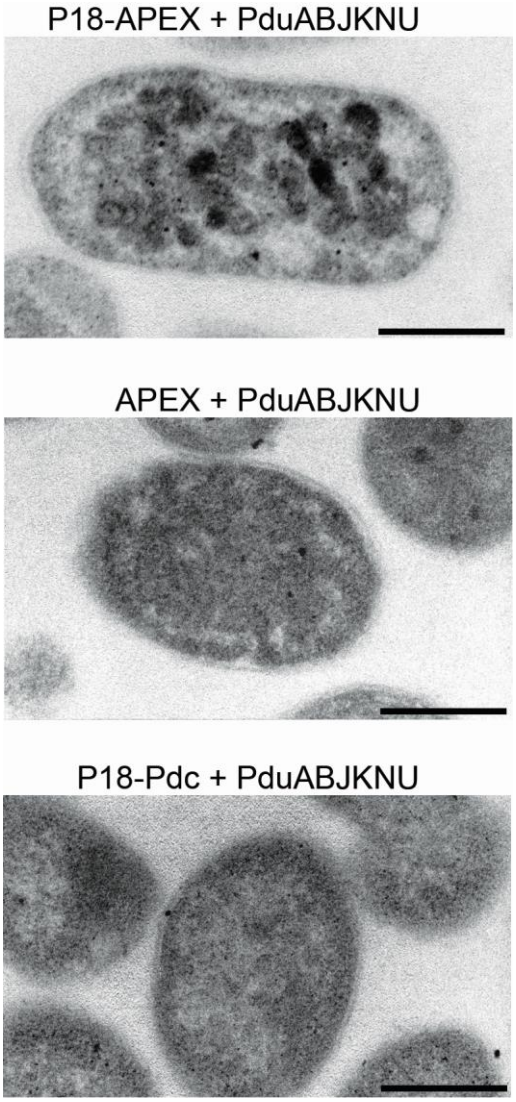


Figure 5.

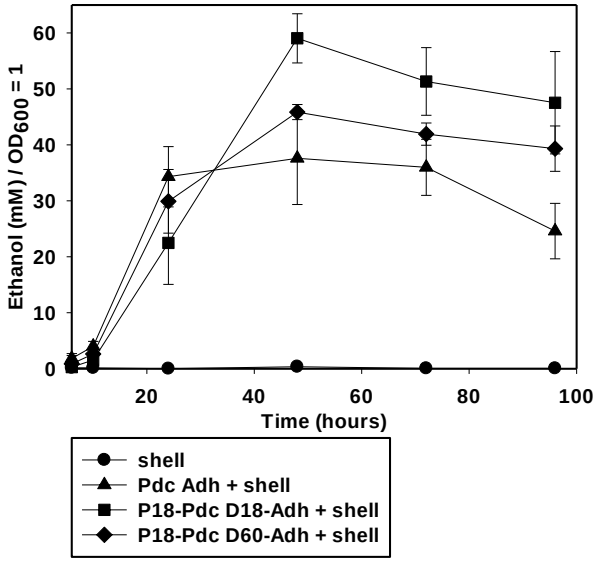

Ethanol (mM) / OD600 = 1
60
50
40
30
20
10
0
20
40
60
80
100
Time (hours)
shell
Pdc Adh + shell
P18-Pdc D18-Adh + shell
P18-Pdc D60-Adh + shell

Figure 6.

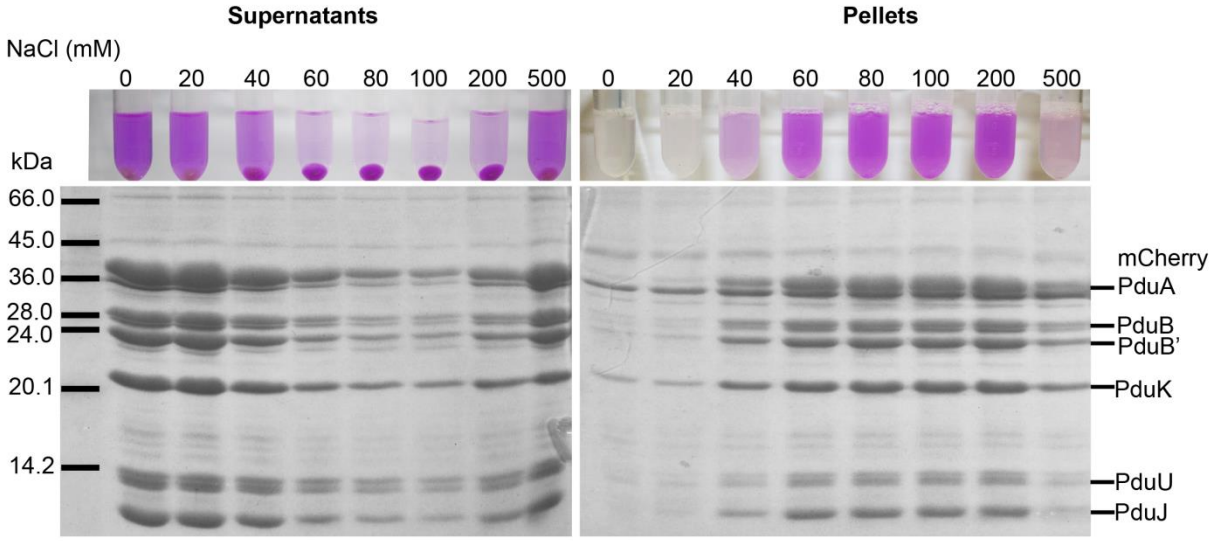
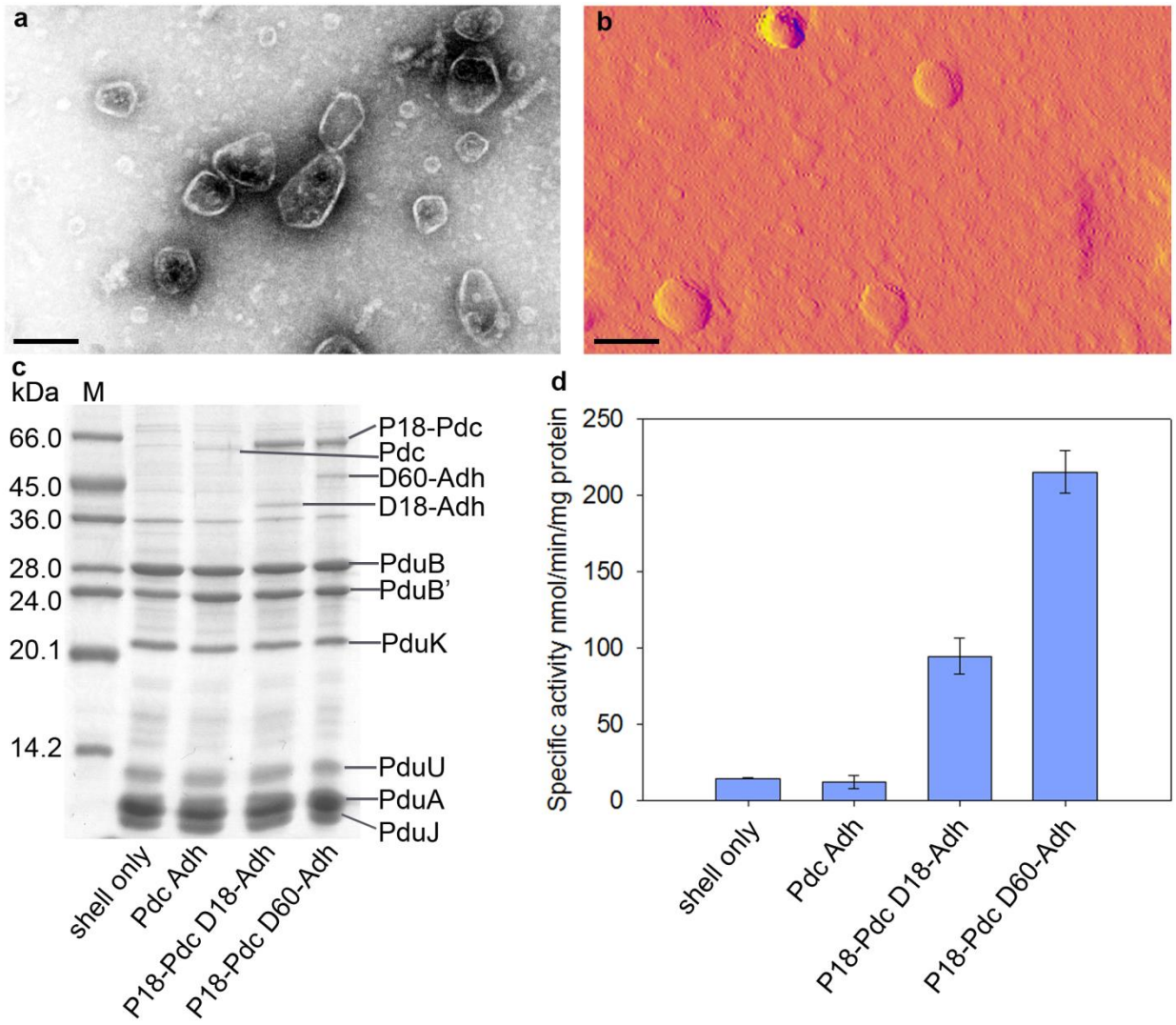

Supernatants
Pellets
NaCl (mM)
0 20 40 60 80 100 200 500
0 20 40 60 80 100 200 500
kDa
66.0
45.0
36.0
28.0
24.0
20.1
14.2
mCherry
PduA
PduB
PduB'
PduK
PduU
PduJ

Figure 7.


a
b
c
kDa M
66.0
45.0
36.0
28.0
24.0
20.1
14.2
P18-Pdc
Pdc
D60-Adh
D18-Adh
PduB
PduB'
PduK
PduU
PduA
PduJ
shell only
Pdc Adh
P18-Pdc D18-Adh
P18-Pdc D60-Adh
d
Specific activity nmol/min/mg protein
250
200
150
100
50
0
shell only
Pdc Adh
P18-Pdc D18-Adh
P18-Pdc D60-Adh

For Table of Contents Only